UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
BOOKSTACKS

Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

http://www.archive.org/details/maximumlikelihoo378atki

Faculty Working Papers

A MAXIMUM LIKELIHOOD APPROACH TO MARKOV MODELS OF BRAND CHOICE

Robert M. Atkinson, II

#378

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

February 22, 1977

A MAXIMUM LIKELIHOOD APPROACH TO MARKOV MODELS OF BRAND CHOICE

Robert M. Atkinson, II

#378

A MAXIMUM LIKELIHOOD APPROACH TO MARKOV MODELS OF BRAND CHOICE

Robert M. Atkinson, II

University of Illinois

December 1976

Not to be quoted without the permission of the author

ABSTRACT

This paper describes a maximum likelihood procedure for estimating the parameters of Markov models of brand choice. For each individual the likelihood function is established for several Markov models based on that individual's purchase history. Those models are compared using the likelihood ratio statistic to determine which model is the best representation. The paper concludes with several suggestions for both empirical and theoretical research.

A MAXIMUM LIKELIHOOD APPROACH TO MARKOV MODELS OF BRAND CHOICE

INTRODUCTION

A Markov model of brand choice may be concerned with either individual or aggregate behavior. Aggregation is a legitimate means of summarizing consumer behavior whenever groups of individuals exhibit some characteristic(s) of homogeneous behavior. Those individuals who have no behavior similar to others may be grouped into a single segment for which no single model applies. Aggregate models may be constructed by modeling the population directly. When aggregate models are constructed by combining individual models, the model that represents the individual will be essentially the same as the population model. Assuming individuals are not completely homogeneous in their behavior, consumers with similar behavior are aggregated into groups. Then it is possible to use a single model to represent almost every subgroup.

It is the purpose of this work to model individual brand choice using Markov models. For each individual's purchase history the parameters of several Markov models are estimated. The models are compared by statistically testing to see which model is the better representation of the individual's purchase behavior. The maximum likelihood method of estimation offers a procedure which focuses on individual purchase behavior.

Anderson and Goodman [1] developed for a homogeneous population (1) maximum likelihood estimates for the transition probabilities, (2) the asymptotic distribution for these estimators, and (3) likelihood ratio an chi-square tests to test the identity, order, and stationarity of the matrix. Massy, Montgomery, and Morrison [9] and Blattberg and Sen [4] mention the usefulness of the maximum likelihood method, but use minimum chi-square techniques.

Literature Review

There are advantages to individual modeling; however, that view is not espoused by everyone. Bass [3] reports the following:

> It has sometimes been argued that if we are to understand consumer behavior well enough at the macro level to construct realistic macro models, we must first understand fully the micro level behavior of consumers. Clearly, this is false. Since there are severe limits to our ability to understand individual behavior, micro models and theories are of very limited value in the construction of macro models. Micro models, on the other hand, are far more useful for managerial purposes of selecting values of decision variables and otherwise determining short-run strategy than micro models, since the variables dealt with (market share, advertising levels, dealing levels, competitive variables) are precisely the variables of managerial relevance. Furthermore, it is far easier to understand macro level behavior of consumers than micro level behavior, since the randomness which characterizes individual behavior tends to be washed out by aggregation.

In support of individual modeling Sheth [10] says:

> Yet another important consideration is that existing stochastic models have so far provided measures of aggregate brand loyalty for the whole market. What we need is a model that will, besides aggregate brand loyal measures, provide measures of each consumer's brand loyalty.

Kotler [6] sees the following benefits from individual modeling:

> 1. These models will force a more rigorous statement of the variables operating in buying situations and the nature of their relationships.
>
> 2. These models will facilitate the comparison of different theories of buyer behavior. The similarities will show up more clearly and the differences can be subjected to sharper empirical tests. The models will highlight the degree to which the differences are structural, functional, or only parametric.
>
> 3. These models will help pinpoint more accurately the data needs of marketing decision makers who seek to understand and to some degree influence consumer behavior.
>
> 4. These models will stimulate inferences not ordinarily obvious from verbal formulations, much in the way a set of postulates about points and lines stimulates a large number of theorems about spatial relations.
>
> 5. These models will facilitate the construction of microconsumer market simulators. The use of microconsuming units in a simulation permits the study of some important properties overlooked in gross market response models, such as the distribution of preferences, the distribution of delays in response to marketing stimuli, and so forth.

Aggregate modeling establishes an average model for the entire population. A minority of unique individuals may be sufficiently different from the majority to reject an aggregate model which accurately portrays most of the population. On the

other hand there may be groups of individuals who behave according to different processes. Research on the individual level presents an opportunity to more explicitly define and model consumer behavior. This decomposition can offer insights and explanations that previously were left untouched.

Rao [9] states that maximum likelihood estimation can be used in place of minimum chi-square estimation. The maximum likelihood technique is applicable in more situations and is superior (on the basis of second-order efficiency and under suitable regularity conditions) to minimum chi-square and associated methods.

THE MAXIMUM LIKELIHOOD METHOD

The likelihood function is the joint density function for obtaining a specific sample given a value for each parameter. The function can also be interpreted as a function of the parameters given the sample. The maximum likelihood method selects the set of parameter estimates which jointly maximize the value of the likelihood function. In other words a parameter value is preferred that implies a large probability of drawing the observed sample. The solution is called a maximum likelihood estimator. This assumes that there are no boundary constraints on the set of admissible parameter values. Any solution which is independent of the observations, i.e., a constant, is ignored.

The maximum likelihood method produces an estimator which is consistent, asymptotically normally distributed, and efficient given certain regularity conditions. All of these desirable properties exist when the likelihood function can be maximized without restriction. When there are constraints on the parameters of the likelihood function, there is no guarantee that all the desirable properties of the maximum likelihood estimator still exist. However, it is not unreasonable to argue that the constrained likelihood function has the same characteristics as any other likelihood function. Atkinson [2] has shown the estimates are consistent and asymptotically unbiased.

Rao establishes a methodology for employing the principle of maximum likelihood. These methods are directly applied to the solution of the brand choice problem. $\hat{p}$ is the maximum likelihood estimate of p, an individual's true purchase probability, such that $L(\hat{p}|x) = \sup L(\hat{p}|x)$ where x represents the individual's purchase history. If the supremum can not be obtained, it may be possible to obtain a near maximum likelihood estimate p^* such that $L(p^*|x) = c \sup L(p|x)$ where c is a fixed number such that $0.0 < c < 1.0$. In other words a near maximum likelihood estimate provides a likelihood value which is some proportion of the maximum likelihood value. Rao proves that a near maximum likelihood estimate, p^*, asymptotically approaches the true maximum likelihood estimate, p, with probability one, without any assumption (restriction) on the admissible set of

estimates. There may be a set of samples for which $\hat{p}$ or p^* does not exist. Under regularity conditions on the joint density of x and p, the frequency of such samples is negligible. In practice it is convenient to work with $\ln L(p|x)$ in which case $\hat{p}$ satisfies the equation $L(\hat{p}|x) = \sup L(p|x)$. When the supremum is attained within the interior of the p-space and $\ln L(p|x)$ is a differentiable function of p, $\hat{p}$ is a solution to the equations

$$\frac{\partial L(p|x)}{\partial p_i} = 0 \quad i = 1,\ldots,q.$$

LIKELIHOOD FUNCTIONS FOR MARKOV PROCESSES

In order to use the maximum likelihood method, the likelihood function must be properly specified for the transitions in the brand choice process. Given a sample of observations $x = (x_1,\ldots,x_n)$, let there be a transition matrix where p_{ij} is the probability of going from state$_i$ to state$_j$, and r_{ij} represents the number of times there is a transition from state$_i$ to state$_j$. $\phi(u,n,J)$ is the set of all possible transition counts which occur from a sample of n consecutive transitions in a Markov chain that has transition matrix J and elements p_{ij}, initial state u, and final state v. The (n by n) matrix of transitions, Z, with elements r_{ij} and range set $\phi(u,n,J)$ has a Whittle Distribution with parameters u, n, and J. (Martin [7])

$$f(\Sigma|u,n,J) = F^{*}_{vu} \frac{\prod_{i=1}^{n} f_{i.}}{\prod_{i=1}^{n}\prod_{j=1}^{n} f_{ij}} p_{ij}^{f_{ij}} \quad 0<p<1$$

$$= 0 \quad \text{otherwise}$$

Σ is an n by n matrix with elements

$$f_{ij} = \delta_{ij} - (f_{ij}/f_{i.}) \text{ for } f_{ij} > 0$$

$$= \delta_{ij} \text{ for } f_{i.} = 0$$

$$f_{i.} = \sum_{j=1}^{n} f_{ij}$$

F^{*}_{vu} is the vu^{th} cofactor of Σ and $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when i is not equal to j. This is the likelihood function for the transition counts of a Markov model; however, to include the entire sample, the first observation must also be accounted for. This is accomplished in a Whittle-1 distribution. (Martin [7])

$$f_{wi}(n,\Sigma|J,u,p) = p_u f_w(\Sigma|u,n,J) \quad 0<p<1$$

$$= 0 \quad \text{otherwise}$$

p_u is a stochastic row vector which is the probability of the first state. Assuming the process begins from a steady-state condition, then p_u is the steady-state probabilities for an ergodic Markov model. In an individual analysis there is a p as well as a p_u associated with each individual. Excluding the proportionality constants, the likelihood functions, maximization equations, and steady-state equations for the Bernoulli, Brand Loyal, and General Markov Models are as follows for an

individual.

---

insert Figure 1 about here

CONSTRAINED MAXIMUM LIKELIHOOD ESTIMATION

Under normal circumstances the range of a parameter is the real line, therefore, there are no restrictions or constraints on the maximum likelihood estimation procedure. In a Markov process the transition probabilities are defined to be between 0.0 and 1.0. Geometrically, the solution space for the Bernoulli model is a plane in the first quadrant of L,p-space that is 1-unit wide with infinite length. The solution space for both the Brand Loyal model and General Markov model is a rectangular parallelopiped in the positive orthant of L,p,k and L,p,q-space with a unit square base and infinite height.

Bernoulli Model

There is a closed form solution for an estimator for any individual purchase probability in the Bernoulli model. No search procedure is necessary; each estimate is a function of the first purchase and the observed transistions.

$$\hat{p} = \frac{f_{11} + f_{01} + u}{\Sigma f + 1}$$

Brand Loyal Model

The partial derivatives of the likelihood function with respect to p and k are quite different. The former is on an individual basis while the latter applies to the entire population. Eliminating fractions the partial with respect to p turns out to be a cubic equation in p.

$$\frac{d\ln L}{dp} = \frac{A}{p} - \frac{B}{1-p} - \frac{Ck}{1-kp} + \frac{1-k}{1-p+kp}$$

$$\frac{d\ln L}{dp} = \Sigma f(k^2-k)p^3+[A(-k^2-k+1)+B+C(2k-k^2)+k^2-1]p^2+(-k-2A-B-Ck+1)p+A=0$$

where $A = f_{01} + f_{11} + u$; $B = f_{10} + 1 - u$; $C = f_{00}$ $\Sigma = f_{00} + f_{01} + f_{10} + f_{11}$. It should be noted that dividing by the coefficient of the p-cubed term, the direction of the cubic is changed if that term is negative; however, the position of the roots is unaffected. (For a closed-form solution to a cubic see Standard Mathematical Tables [11])

A cubic equation

$$y^3 + Py^2 + Qy + R = 0$$

may be reduced to the form

$$x^3 + Ax + B = 0$$

by substituting

$$y = x - (P/3).$$

After the substitution, the coefficients in the new cubic equation have the following relationship with the original coefficients.

$$A = (3Q - P^2)/3; \quad B = (2P^3 - 9PQ + 27R)/27$$

The solutions to the reduced form cubic are

$$X = (R+S,\ -\frac{R+S}{2} + \frac{R-S}{2}\sqrt{-3},\ -\frac{R+S}{2} - \frac{R-S}{2}\sqrt{-3})$$

where $R = \left[-\frac{B}{2} + \sqrt{\frac{B^2}{4} + \frac{A^3}{27}}\right]^{1/3}$ $S = \left[-\frac{B}{2} - \sqrt{\frac{B^2}{4} + \frac{A^3}{27}}\right]^{1/3}$

Examining the term in B and 3 under the square root sign yields information about the structure of the three roots.

$(B^2/4) + (A^3/27) < 0$ there are 3 real unequal roots

$= 0$ there are 3 roots and at least 2 are equal

> 0 there is 1 real root and 2 complex conjugate roots

In the last case a trigonometric solution is necessary.

$$\cos(\phi) = (-B/2)/(-A^3/27)^{1/2}$$

Now the solutions are

$$x = 2(-A/3)^{1/2}\cos(\phi/3)$$

$$x = 2(-A/3)^{1/2}\cos(\phi/3 + 120°)$$

$$x = 2(-A/3)^{1/2}\cos(\phi/3 + 240°)$$

The equation to be solved in the Brand Loyal model has the form

$$p^3 + R_1 p^2 + R_2 p + R_3 = 0$$

$$R_1 = \frac{A(-k^2+k+1)+B+C(2k-k^2)+k^2-1}{\Sigma f\ (k^2-k)}$$

$$R_2 = \frac{-2A-B-Ck+1-k}{\Sigma f(k^2-k)} \qquad R_3 = \frac{A}{\Sigma f(k^2-k)}$$

The solution requires a grid search on k. For a given k a set of p's is calculated from the cubic equation. The set of p's and the corresponding k that produce the largest likelihood value in the required unit-interval are the solution.

There are two special cases associated with the solution of MLE.

CASE I: $R_3 = 0$

This particular case occurs whenever the purchase sequence consists of all zeros. For any purchase length of all zeros, A = 0; B = 1; C = some value, e.g., C* and $\Sigma f = A + B + C - 1 = C^*$. If A, B, and C are substituted into the original cubic equation, the new equation is

$$[C^*(2k-k^2)+k^2]p^2 - (C^*k+k)p = 0.$$

Obviously with $R_3 = 0$, p = 0 is a root and it can be shown that the other two roots are real and outside the 0-1 interval.

CASE II: The purchase sequence is all ones

When the purchase sequence is all 1's, A, B, and C are B = C = 0 and A = some value, e.g., A* and $\Sigma f = A + B + C - 1 = A^* - 1$. Substituting these values into the original cubic equation, the new equation is

$$[(A^*-1)(k^2-k)]p^3+[A^*(-k^2+k-1)+B+C(2k-k^2)+k^2-1]p^2+(-2A^*+1-k)p+A=0$$

It can be shown that regardless of the number of ones in the sequence $p = 1$ is always a root, i.e., division with no remainder. It can also be shown that the other two roots are always greater than 1. Therefore, $p = 1$ is the only root and can be identified immediately whenever $A = A^*$ and $B = C = 0$. It can be intuitively argued that there is always a root between zero and one. The two extreme cases (i.e., all ones or all zeros) result in p's equal to zero and one, respectively. Any other sequence is a combination of the two extremes and therefore should have at least one root such that $0 < \text{root} < 1$. This author was unable to prove mathematically the above result. However the computer search is designed to stop if there is no root between zero and one.

Individual Brand Loyal Model

The Individual Brand Loyal model is a modification of Morrison's Brand Loyal model. Fundamentally they are identical except in the Individual Brand Loyal model each individual has a k, i.e., k is not a population parameter but an individual parameter. There is a partial derivative with respect to k for each individual. The Morrison Brand Loyal model has $n+1$ parameters while the Individual Brand Loyal has $2n$. In terms of

steady-state equations and maximum likelihood estimation, the solution process is virtually the same. The only difference between this model and the General Markov model is that k must be less than one in the Individual Brand Loyal model.

General Markov Model

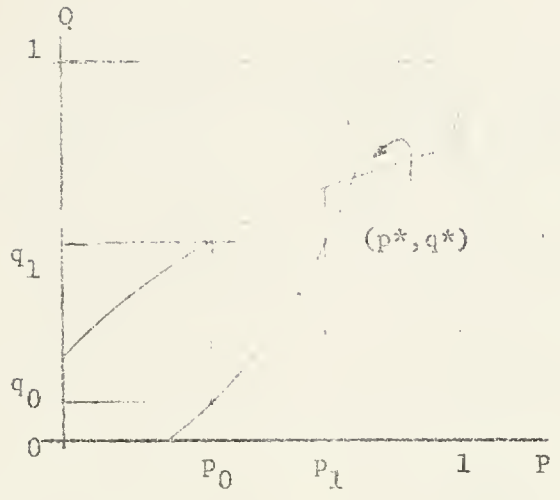
The likelihood function for the General Markov model is a concave function and consequently a maximum exists. (See Appendix A) In the General Markov model there is a derivative with respect to p and q for each individual.

$$\frac{\partial \ln L}{\partial p} = \frac{-f_{00}}{1-p} + \frac{f_{01} + u}{p} - \frac{1}{p+q} = 0 \qquad \frac{\partial \ln L}{\partial q} = \frac{f_{10} + 1 - n}{q} - \frac{f_{11}}{1-q} - \frac{1}{p+q} = 0$$

Simultaneous solution of the equations show that p is a quadratic function of q and vice-versa. It is possible to solve these equations simultaneously and obtain a closed form solution for p and q. However, that analytic solution ignores the fact that solution p and q's are restricted to the unit interval.

Another method of solution which applies only for the General Markov model is illustrated in Figure 2. Projecting the likelihood onto the p,q-plane graphically portrays solutions.

Figure 2



PROJECTION OF THE MAXIMUM LIKELIHOOD SOLUTION

The intersection is the point of maximum likelihood and may occur at a boundary. p may be any combination of "0" or "1" (boundary solutions) or quadratic for various values of q and vice-versa. Any given starting point q_o will generate a p_o, which in turn generates a q_1. The iteration continues increasing the value of the likelihood function until reaching p* and q*. The same result will be achieved if p_o is the given starting point and q's are solved for. Since a concave function is also unimodal, the solutions are computed using a golden section search technique in two variables. The details of this search method are expressed in Bauer and Thomson [5].

APPLICATION

Applying the maximum likelihood method to panel data requires the following steps. The brands chosen by each individual must be recoded such that the brand under investigation is a one and all others are zero. This simplifies the analysis; however, there is a loss of information in this reduction process. If the individual purchase histories have different lengths, it does not affect the analysis. Each purchase history is considered to be independent from all others. Since each history is analyzed separately, there is no need for them to have the same length. The first and last observation and transition counts are appropriately placed in the Whittle-1 Distribution. This establishes the likelihood function for the transition counts in a Markov process. Depending on the model, an analytical method or search technique is used to estimate the purchase probabilities. For each individual the parameters of several Markov models are estimated. In order to ascertain the most parsimonious representation of the purchase behavior from the set of candidate models, the models are tested to see if they are significantly different. Once a model is chosen the estimate of an individual's probability of purchase and type of purchase process based on the purchase history is complete.

LIKELIHOOD RATIO TEST

Using the maximum likelihood estimators, the maximum value of the likelihood function can be used to test hypotheses. If the null hypothesis is concerned with a model that has a parameter vector p_o, the likelihood function value is $L(x|\hat{p}_o)$. For the alternative hypothesis another model has parameter vector p_1 with likelihood value $L(x|\hat{p}_1)$. x in both cases is an individual purchase history. The likelihood ratio is

$$\lambda = \frac{L(x|\hat{p}_o)}{L(x|\hat{p}_1)}$$

If λ $\begin{cases} > 1.0, \text{ reject the null hypothesis} \\ < 1.0, \text{ accept the null hypothesis} \\ = 1.0, \text{ there is no definite conclusion} \end{cases}$

In order to test the significance of the difference between two models, it is necessary to see whether lambda is significantly different from one. For large samples $-2 \ln \lambda$ is distributed chi-square with $q - r$ degrees of freedom. q is the number of parameters in the null hypothesis model and r the number of parameters in the alternative hypothesis model. The likelihood ratio test is used to determine which Brand choice model better describes the purchasing behavior of populations and individuals.

The Morrison Brand Loyal model is not considered as one of the models to represent individual purchase behavior. A single k for all consumers may weaken the assumption of independent

purchase histories. The Individual Brand Loyal model is a modification of the Morrison Brand Loyal model which addresses individual purchase behavior. However, the Individual Brand Loyal model is a combination of the Bernoulli and General Markov models. When K is greater than one, the estimate is set equal to one and the Individual Brand Loyal model becomes a Bernoulli model. If K is less than one, the range of K is within the acceptable range for the model.

The purchase sequence 01001 produces the following results.

Model	Estimate(p,K)	Likelihood
Bernoulli	(0.400, 1.00)	0.06912
General Markov	(0.001, 579.00)	0.18007

Likelihood Ratio Statistic	1.915
Type-I Error (alpha)	0.2
Chi-square Critical Value	1.642

Choose the General Markov model

The null hypothesis and numerator of the likelihood ratio test is the model with the fewest parameters. For this example the null hypothesis is the Bernoulli model and the alternative hypothesis is the General Markov model. If the critical value of the chi-square test is greater than the likelihood ratio statistic, the null hypothesis is accepted. For the purchase history 01001 the likelihood values for both models are compared, i.e., $-2\ln(.069/.18)=1.915$. Since the likelihood ratio statistic exceeds the table value, the General Markov model is selected as the better representation.

The purchase sequence C0011C produces the following results.

Model	Estimate(p,k)	Likelihood
Bernoulli	(0.333, 1.0)	0.02195
General Markov	(0.240, 0.724)	0.04355

Likelihood Ratio Statistic	1.371
Type-I Error (alpha)	0.20
Chi-square Critical Value.	1.642

Choose the Bernoulli model

Since the likelihood ratio statistic is less than the chi-square critical value, the null hypothesis is accepted thereby selecting the Bernoulli model.

CONCLUSIONS

The objective of this work was to develop the maximum likelihood approach for estimating the parameters of Markov models of consumer behavior and to use the likelihood ratio statistics for classifying individuals by the model which best represents their observed behavior. For the practicing marketer this approach provides a framework for attacking several important problems. Classification of consumers by their purchase behavior naturally implies market segmentation. Market segments can be combined to develop market shares. The market share and segments developed from the maximum likelihood method can be applied to panel data used in other studies. An empirical comparison can be made to see and explain similarities and differences. Also, market segmentation can be divided into several areas. Since the maximum likelihood method specifically

concerns itself with each individual consumer, aggregation into segments can occur along different dimensions. Consumers can be grouped by their response to either changes in the total marketing mix or any of its individual components. Response can be measured as a significant change in purchase probabilities. If attention is focused on the product and promotion phase of the marketing mix, the results have an impact on the evaluation of product and promotion management. For example, the consumer responses or changes in the response to the product could influence decisions to introduce a new product, modify an existing product, or eliminate a product.

The maximum likelihood method also opens the door for further empirical research in related areas. It may be possible to establish a relationship between the probability of purchase and the marketing mix or any combination of its components. For example, the probability of purchase is modeled as some function of the product's price and promotion. A relationship may be established between the probability of purchase and the demographics of the consumers. For example, the probability of purchase is modeled as some function of the consumers' ages and incomes. These first two suggestions could be combined such that the probability of purchase is a function of marketing and demographic variables.

The methodology employed for the solution of a two-state model can be expanded to an n-state model. Theoretically, there appear to be few problems. Practically, there may be problems associated with collecting sufficient data to estimate all the parameters. It may be possible to apply the maximum likelihood method to other models of consumer behavior. If the likelihood function for a model can be established, then the methodology employed in this paper sets up a possible framework for solution. In addition, the maximum likelihood approach allows an analysis to be performed as either several models across consumers or consumers across given models.

APPENDIX A

The likelihood function for the General Markov model is concave and consequently a maximum exists.

Eqs. (1) and (2) are the gradient of the likelihood function.

$$\frac{\partial \ln L}{\partial p} = \frac{-f_{00}}{1-p} + \frac{f_{01}+u}{p} - \frac{1}{p+q} \quad (1) \qquad \frac{\partial \ln L}{\partial q} = \frac{f_{10}+1-u}{q} - \frac{f_{11}}{1-q} - \frac{1}{p+q} \quad (2)$$

Concavity is defined as follows for two variables.

$$\frac{\partial^2 f(x,y)}{\partial x^2} \cdot \frac{\partial^2 f(x,y)}{\partial y^2} - \left[\frac{\partial^2 f(x,y)}{\partial x \partial y}\right]^2 \geq 0 \quad (3)$$

$$\frac{\partial^2 f(x,y)}{\partial x^2} \leq 0 \quad (4) \qquad \frac{\partial^2 f(x,y)}{\partial y^2} \leq 0 \quad (5)$$

If these conditions are satisfied the function is concave. Eq. (3) may be rewritten as follows.

$$\frac{\partial^2 f(x,y)}{\partial x^2} \cdot \frac{\partial^2 f(x,y)}{\partial y^2} = \left[\frac{\partial^2 f(x,y)}{\partial x \partial y}\right]^2 \quad (6)$$

Eqs. (7), (8), and (9) are the higher order derivatives for the General Markov likelihood function necessary to test for concavity.

$$\frac{\partial^2 \ln L}{\partial p^2} = \frac{f_{00}}{(1-p)^2} - \frac{f_{01}+u}{p^2} + \frac{1}{(p+q)^2} \quad (7)$$

$$\frac{\partial^2 \ln L}{\partial q^2} = \frac{-f_{11}}{(1-q)^2} - \frac{f_{10}+1-u}{q^2} + \frac{1}{(p+q)^2} \quad (8)$$

$$\frac{\partial^2 \ln L}{\partial p \partial q} = \frac{1}{(p+q)^2} \quad (9)$$

In order to satisfy eq. (4), the following relationship must be true.

$$\frac{1}{(p+q)^2} - \frac{f_{00}}{(1-p)^2} - \frac{f_{01} + u}{p^2} \leq 0$$

$$(1-p)^2 p^2 \leq (p+q)^2 [p^2 f_{00} + (1-p)^2 (f_{01} + u)]$$

u is the first observation.

f_{00} is the number of "00" transitions
f_{01} is the number of "01" transitions
f_{10} is the number of "10" transitions
f_{11} is the number of "11" transitions

For any single transition the quantity $f_{00} + f_{01} + u = 1$. For more transitions than one $f_{00} + f_{01} + u > 1$. For only one transition either $f_{00} = 0$ or $f_{01} + u = 0$. If $f_{00} = 0$,

$$\frac{1}{(p+q)^2} < \frac{f_{01} + u}{p^2}$$

If $f_{01} + u = 0$, the estimate of p is essentially equal to zero and since u = 0, the estimate for $\hat{p} = 1$.

$$\frac{1}{(p+q)^2} \leq \frac{f_{00}}{(1-p)^2}$$

$$1 \leq f_{00}$$

Therefore, the conditions set forth in eq.(4) are satisfied.

Using the previous line of reasoning, the conditions set forth in eq. (5) are also satisfied.

$$\frac{1}{(p+q)^2} - \frac{f_{11}}{(1-q)^2} - \frac{f_{10} + 1 - u}{q^2} \leq 0$$

$$(1-q)^2 q^2 \leq (p+q)^2 [q^2 f_{11} + (1-q)^2 (f_{10} + 1 - u)]$$

Let the value of eqs. (7), (8), and (9) be

$$-A+x,\ -B+x, \text{ and } x \text{ where } x = 1/(p+q)^2 .$$

The conditions set forth in eq. (3) are

$$(-A+x)(-B+x) \geq x^2$$

$$AB - (A+B)x + x^2 \geq x^2$$

$$AB - (A+B)x \geq 0$$

$$AB \geq (A+B)x$$

For a single transition A and B have the following values.

$A = -1/(1-p)^2$ for a "00" and $-1/p^2$ for all others

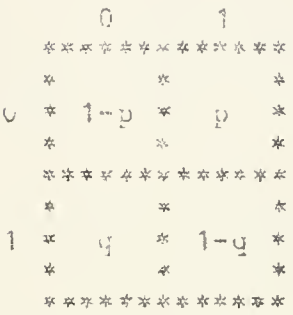
$B = -1/(1-q)^2$ for a "11" and $-1/q^2$ for all others

For all cases AB is positive while A + B is negative; therefore, it is true that $AB \geq (A + B)x$. Having satisfied all the conditions, the likelihood function for the General Markov model is concave and a maximum exists.

APPENDIX B

THE LIKELIHOOD SOLUTION OF THE GENERAL MARKOV MODEL

The General Markov model has the following transition matrix.



The steady-state solution of this matrix follows.

$$p(0) = \frac{q}{p+q} \qquad p(1) = \frac{p}{p+q}$$

The likelihood density function for a Markov process is the Whittle-1 distribution. If the constant of proportionality is ignored, the likelihood function may be expressed as proportion.

$$L \propto (1-p_i)^{f_{00}} p_i^{f_{01}} (1-q_i)^{f_{11}} q_i^{f_{10}} q_i^{1-u} p_i^{u} / (p_i + q_i)$$

u is the first observation. It is assumed that the first observation occurs after the consumer's purchase behavior has reached a steady-state condition. The natural logarithm of the likelihood function follows.

$$\ln L \propto f_{00}\ln(1-p_i)+f_{01}\ln p_i+f_{11}\ln(1-q_i)+f_{10}\ln q_i+(1-u)\ln q_i$$

$$+u\ln p_i-\ln(1+p_i)$$

The natural log of the likelihood function may be differentiated with respect to the two parameters of the model, p and q, for each individual. If the i-subscripts are dropped, the maximization equations with respect to p and q for the i^{th} individual follow.

$$\frac{\partial \ln L}{\partial p} = \frac{f_{00}}{1-p} + \frac{f_{01}}{p} + \frac{u}{p} - \frac{1}{p+q} = \frac{f_{01}+u}{p} - \frac{f_{00}}{1-p} - \frac{1}{p+q} \qquad (1)$$

$$\frac{\partial \ln L}{\partial q} = \frac{-f_{11}}{1-q} + \frac{f_{10}}{q} + \frac{1-u}{q} - \frac{1}{p+q} = \frac{f_{10}+1-u}{q} - \frac{f_{11}}{1-q} - \frac{1}{p+q} \qquad (2)$$

The simplification of eq. (1) follows.

$$\frac{(f_{01}+u)(1-p) - f_{00}p}{p(1-p)} = \frac{1}{p+q}$$

$$\frac{p(1-p)}{(f_{01}+u)(1-p) - f_{00}p} = p+q$$

Let $A = f_{00}$ and $B = f_{01}+u$.

$$-A(p+q)+B(1-p)(p+q)-p(1-p) = 0$$

$$-A(p^2+qp)+B(p+q-p^2-qp)-p+p^2 = 0$$

$$(-A-B+1)p^2+B(-A+B-Bq-1)p+Bq = 0$$

Using the quadratic formula, p can be solved for in terms of q.

$$p = \frac{-(B-1-Aq-Bq) \pm [(B-1-Aq-Bq)^2 - 4(-A-B+1)Bq]^{\frac{1}{2}}}{2(-C-D+1)} \qquad (3)$$

The fact that p is quadratic in q and q is quadratic in p agrees with the proof in Appendix A. The likelihood function for the General Markov model is concave.

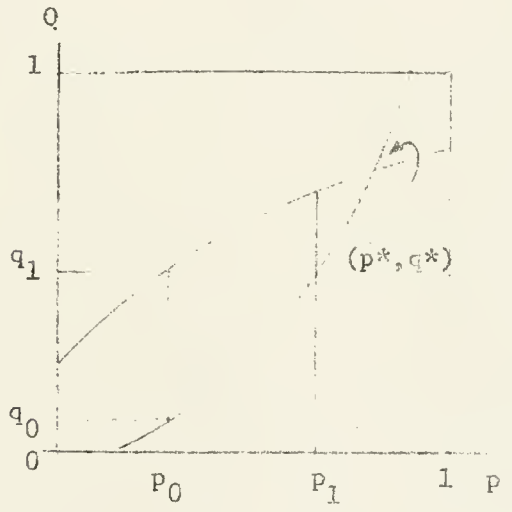
Since p and q are probabilities, they have a range such that $0 < p, q < 1$. However, in the closed form solution it is possible that $q > 1$, i.e., $(-C-D+1) = 0$. The constraints on q are acknowledged by setting $q = 1$ and solving for p.

$$p = \frac{-(-A-1) \pm [(-A-1)^2 - 4(-A-B+1)B]^{1/2}}{2(-A-B+1)}$$

The same conditions hold for a solution to q when $p = 1$.

$$q = \frac{-(-D-1) \pm [(-D-1)^2 - 4(-C-D+1)C]^{1/2}}{2(-C-D+1)}$$

p and q may be plotted in terms of each other.



Usually, the solution or intersection is within the zero to one limits. The intersection is the point of maximum likelihood and may occur at a boundary. The solution set for p is one of the following:

$$p \begin{cases} = \text{boundary solution (zero for all } q) \\ = \text{quadratic solution in } q \\ = \text{boundary solution (one for all } q) \end{cases}$$

Any starting value for q such as q_0 will generate a p_0 and p_0 generates a q_1 until the maximum likelihood point of intersection or a boundary is reached. In other words successive iterations converge. In order to account for the boundary solutions, i.e., the constraints on the probabilities, a golden section search is used rather than a closed-form solution of the quadratic equations.

REFERENCES

1. Anderson, T.W. and Leo A. Goodman, Statistical Inference About Markov Chains, ANNAL OF MATHEMATICAL STATISTICS, Vol. 28, 1957, pp. 89-109.

2. Atkinson, Robert A., II, A Maximum Likelihood Approach To Stochastic Models Of Brand Choice, unpublished. Ph.D. lissertation at the Graduate School Of Industrial Administration, Carnegie-Mellon University, 1976.

3. Bass, Frank, The Theory Of Stochastic Preference And Brand Switching, JOURNAL OF MARKETING RESEARCH, Vol. 11, 1974, pp. 1-20.

4. Blattberg, Robert and Subrata Sen, An Evaluation Of The Application Of Minimum Chi-square Procedures To Stochastic Models Of Brand Choice, JOURNAL OF MARKETING RESEARCH, 1973, pp. 421-427.

5. Gaver, Donald P. and Gerald L. Thompson, PROGRAMMING AND PROBABILITY MODELS IN OPERATIONS RESEARCH, Brooks/Cole Publishing Co., Monterey, Calif., 1973, pp. 252-255

6. Kotler, Philip, Mathematical Models Of Individual Buyer Behavior, BEHAVIORAL SCIENCE, 13, 1968, pp. 274-287.

7. Martin, J.J., BAYESIAN DECISION PROBLEMS AND MARKOV CHAINS, John Wiley and Sons, Inc., New York, 1967.

8. Massy, William F., David B. Montgomery, and Donald G. Morrison, STOCHASTIC MODELS OF BUYING BEHAVIOR, The M.I.T Press, Cambridge, Mass., 1970, pp. 1-140.

9. Rao, C. Radhakrishna, LINEAR STATISTICAL INFERENCE AND ITS APPLICATIONS, John Wiley and Sons, Inc., New York, 1965, pp. 283-291.

10. Sheth, Jagdish N., A Factor Analytical Model Of Brand Loyalty, JOURNAL OF MARKETING RESEARCH, 5, 1968, pp. 395-404.

11. Standard Mathematical Tables edited by Samuel Selby, 19th edition, Cleveland, Ohio, p. 103.

FIGURE 1

	Bernoulli	Morrison on Brand Loyal	General Markov
Steady-State probabilities	$p_0 = 1-p$ $p_1 = p$	$p_0 = (1-p)/(1-p+kp)$ $p_1 = kp/(1-p+kp)$	$p_0 = q/(p+q)$ $p_1 = p/(p+q)$
Notes * u is the first purchase for each individual		There is a single k for the entire population	There is a k for each individual; $k = p/(1-q)$
Transition Definitions	$A = f_{01} + f_{11} + u$ $B = f_{00} + f_{10} + 1 - u$	$A = f_{00}$; $B = f_{01} + u$ $C = f_{11}$; $D = f_{10} + 1 - u$	$A = f_{00}$; $B = f_{01} + u$ $C = f_{11}$; $D = f_{10} + 1 - u$
Likelihood function	$L \propto P^{A} (1-p)^{B}$	$L \propto (1-kp)^{A} (kp)^{B} p^{C} (1-p)^{D}/(1-p+kp)$	$L \propto (1-p)^{A} p^{B} (1-q)^{C} q^{D}/(p+q)$
Derivatives and/or Solutions	$\frac{\partial \ln L}{\partial p} = \frac{-B}{1-p} + \frac{A}{p}$ $\hat{p} = A/(A+B)$ (analytic solution)	$\frac{\partial \ln L}{\partial p} = \frac{B+C}{p} - \frac{D}{1-p} - \frac{Ak}{1-kp} + \frac{1-k}{1-p+kp}$ $\frac{\partial \ln L}{\partial k} = \sum_i \left[\frac{B_i}{k} - \frac{A_i p_i}{1-kp_i} - \frac{p_i}{1-p_i+kp_i} \right]$ (search procedure)	$\frac{\partial \ln L}{\partial p} = \frac{B}{p} - \frac{A}{1-p} - \frac{1}{p+q}$ $\frac{\partial \ln L}{\partial q} = \frac{C}{q} - \frac{D}{1-q} - \frac{1}{p+q}$ (search procedure)

UNIVERSITY OF ILLINOIS-URBANA
3 0112 060296552